

CORPUS

**Corpus**

4 | 2005

Les corpus politiques : objet, méthode et contenu

---

## Les corpus politiques : objet, méthode et contenu.Introduction

Damon Mayaffre

---



### Édition électronique

URL : <http://journals.openedition.org/corpus/292>

ISSN : 1765-3126

### Éditeur

Bases ; corpus et langage - UMR 6039

### Édition imprimée

Date de publication : 1 décembre 2005

ISSN : 1638-9808

### Référence électronique

Damon Mayaffre, « Les corpus politiques : objet, méthode et contenu.Introduction », *Corpus* [En ligne], 4 | 2005, mis en ligne le 05 septembre 2006, consulté le 04 mai 2019. URL : <http://journals.openedition.org/corpus/292>

---

Ce document a été généré automatiquement le 4 mai 2019.

© Tous droits réservés

---

# Les corpus politiques : objet, méthode et contenu.Introduction

Damon Mayaffre

---

- 1 Le titre du présent numéro de CORPUS est à lui seul critique. Qualifier *corpus* par *politique*, c'est indiquer que le matériel linguistique qui a été rassemblé pour être traité a été choisi pour sa valeur extra-linguistique. Délibérément, c'est poser le corpus comme un lieu problématique de rencontre entre la langue et la société; l'analyste jouant le rôle conscient et déterminant d'entremetteur.
- 2 La question des corpus, de leur constitution, de leur traitement, de leur usage a été posée dès l'origine par les analystes du discours politique français. [Dubois 1969 : 5-7], après sa thèse sur le vocabulaire politique et social, initie ainsi la réflexion dès 1969 dans l'introduction du numéro de *Langages* fondateur de l'Analyse du Discours (désormais AD) en France. Le laboratoire de Saint-Cloud, par la plume de [Peschanski 1984] par exemple, a plusieurs fois estimé indispensable de consacrer des articles aux corpus vus par le prisme contraignant de la lexicométrie politique. D'un point de vue plus épistémologique, [Guilhaumou *et alii* 1994] font du corpus le fil conducteur de leur ouvrage via une réflexion sur l'archive et le discours. Et ici même, dans le premier numéro de CORPUS, [Guilhaumou 2002] a pu retracer l'histoire de la notion dans le champ d'une AD française très marquée par les corpus politiques<sup>1</sup>.
- 3 Finalement, depuis près de 40 ans, aucune étude qui entend appréhender les discours ou les textes dans leur dimension sociale ou politique n'est indifférente à la question des corpus pour la raison suffisante qu'une production linguistique revêt une dimension politique seulement si elle se trouve effectivement attestée dans la société. L'étude des données attestées n'implique certes pas automatiquement ou systématiquement un corpus – comme les données non attestées n'excluent pas que le syntacticien constitue un recueil de phrases fabriquées – mais elle en laisse néanmoins pressentir violemment la nécessité.
- 4 Deux idées fondamentales travaillent les analystes de corpus politiques. Toutes deux semblent mener à la *linguistique de corpus* telle qu'elle se définit depuis quelques années dans le monde anglo-saxon [Biber, Conrad & Reppen 1998 ; Tognini-Bonelli 2001 ; Aijmer

and Altenberg (ed.) 2002] ou en France [Habert, Nazarenko et Salem 1997 ; Rastier 2001, 2005-a, 2005-b].

Pour une linguistique des corpus politiques

- 5 La signification se construit en langue, mais il n'y a de sens – et à *fortiori* de sens politique – qu'en contexte. L'affirmation n'est aujourd'hui plus contestée jusqu'à en devenir banale mais son explicitation continue de déranger. La compréhension littérale du mot *con-texte* semble pourtant ne pouvoir échapper à personne. Sans parler de l'horizon extra-linguistique infini que suggère le préfixe, la focale sur laquelle s'arrête le radical est bien le texte. L'objet pertinent du sémanticien n'est donc pas le mot ou la phrase – « artefact de grammairien » [Rastier 2001 : 30]– mais le texte. Il y a, là, une position théorique que de plus en plus de linguistes osent assumer depuis la redécouverte des œuvres de Hjelmslev et de Bakhtine ([Hjelmslev 1943 (1968-1971)] et [Bakhtine 1952-1953 (1984)] ; on citera dans des perspectives différentes : [Halliday et Hasan 1976], [Van Dijk 1984], [Combettes 1983], [Bronckart 1997], [Rastier 1989 et 2001] ; [Amossy 2002] ; [Adam 1990 et 1999] ; etc. On se reportera nécessairement à ce dernier pour la bibliographie commentée de cette linguistique du texte : [Adam 1999 : *Introduction* : 5-20 et *Chapitre I : Pour une linguistique des grandes unités* : 21-42]).
- 6 De fait, en tous cas, toutes les contributions de ce numéro rassemblent des corpus *textuels*. Il existe bien en analyse du discours politique des expérimentations sur des corpus de phrases, elles-mêmes sélectionnées autour de mots-pivots, mais aucune n'entend déraciner ces phrases ou ces mots de leur (co(n)texte d'origine<sup>2</sup>. Ici même, l'article de Edward Arnold, celui de Dominique Desmarchelier, celui de Jean-Marc Leblanc et Williams Martinez, celui de Denis Monière, Cyril Labbé et Dominique Labbé, qui reprennent de la grande tradition lexicométrique l'approche statistique paradigmatique, prennent soin au moment de l'acte interprétatif de resituer les mots, les co-occurents, les codes grammaticaux dans leur cadre *textuel*, étape préalable indispensable avant de les resituer dans un cadre *con-textuel* élargi<sup>3</sup>. De manière plus problématique, mais finalement identique, Mathieu Brugidou réfléchit dans son article sur l'élaboration de corpus « d'énoncés d'opinion publique ». Ceux-ci, constitués de questions-réponses issues de sondages, seront considérés comme des entités textuelles à part entière qu'il convient d'analyser.
- 7 Pour François Rastier dont l'œuvre constitue le cadre théorique de l'article de Carine Duteil, la réflexion épistémologique est plus avancée. Si le texte est l'unité fondamentale de la linguistique<sup>4</sup>, c'est le corpus qui en est l'unité maximale. La contextualisation – c'est-à-dire, au fond, la recherche du sens – se fait, au-delà du texte, en corpus :
 

« tout texte placé dans un corpus en reçoit des déterminations sémantiques, et modifie potentiellement le sens de chacun des textes qui le composent » [Rastier 2001 : 92].
- 8 Le sens en effet n'est jamais donné. Il est toujours construit lors de parcours interprétatifs complexes que le linguiste est en charge de contrôler et dans lesquels s'articulent écriture et réécriture, textes et contexte, conditions culturelles d'émission, de réception et d'analyse. Précisément, ce contrôle du parcours interprétatif s'effectue autant que possible *dans et par* le corpus, conçu comme la seule ou l'ultime<sup>5</sup> entité possible d'objectivation du contexte pour la linguistique. (A cette fin, nous avons proposé qu'un corpus performant devait être *réflexif* : l'ensemble du corpus intègre et représente le co-texte immédiat, nécessaire à l'interprétation, de chacun des textes constitutifs. *Les corpus réflexifs* peuvent être ainsi considérés comme des tout sémantiques, certes pas idéaux et

exclusifs, mais auto-suffisants, dans lesquels les textes, organisés en réseau signifiant, s'éclairent les uns les autres [Mayaffre 2002-b]).

- 9 Dès lors, en tant que condition d'objectivation d'un contexte nécessaire à l'appréhension du sens, le corpus peut hautement se définir comme *le lieu linguistique où se construit et s'appréhende le sens*. C'est dire s'il n'est pas, pour nous, une fantaisie mais un essentiel.
- 10 Dans la dernière livraison de la revue, [Scheer 2004-a et 2004-b] a montré comment le corpus pouvait être, pour les phonologues, l'observatoire de noumènes linguistiques *a priori*. Pour l'auteur, l'intuition linguistique semble précéder l'objet, comme l'essence l'existence et la langue la parole. Le corpus atteste, valide, illustre une intuition linguistique qui aura présidé à sa constitution. Même en insistant sur la dimension *heuristique* du corpus, l'auteur pose le corpus comme un « lieu théorique » [Scheer 2004-b : 182] en phonologie dans lequel le linguiste cherche ou plutôt « montre » [ibid.] ce qu'il a pressenti. Le propos est convaincant contre une linguistique purement empirique et descriptive qui, en phonologie, prend la forme du tout phonétique. De fait, la tradition anglo-saxonne, dont T. Scheer remonte la généalogie jusqu'à Locke ou Hume a poussé loin le béhaviorisme faisant du corpus un objet froid, suffisant, soit-disant objectif, dont la description serait le commencement et la fin. De fait encore, la course éperdue à des corpus numérisés, toujours plus grands, dont le seul mérite est d'exister, n'a pas donné, malgré les réussites de [Biber 1988 et 1995] aux USA ou de [Brunet 1981] en France, les résultats théoriques linguistiques à la hauteur des investissements consentis en terme de saisie, de codage, d'archivage. C'est l'empirie bornée qui, effectivement, dans les années 70-80 a hypothéqué la linguistique de corpus en prêtant le flanc à une critique chomskienne pourtant irrecevable à force d'être excessive<sup>6</sup>.
- 11 Pour la majeure partie des articles de ce numéro, le propos est assez différent<sup>7</sup> même si tout un chacun se retrouve sur les trois points qui constituent le socle éditorial minimal de CORPUS tel qu'il a été défini lors du premier numéro : i) le caractère nécessaire des corpus en linguistique, ii) la valeur heuristique des corpus plus encore que leur valeur validante, iii) l'importance des hypothèses de travail en amont de la sélection et de l'organisation des données qui font du corpus un objet non pas donné mais construit (voir les différents articles du numéro fondateur [CORPUS 2002], et un rappel dans [Scheer 2004-b]).
- 12 Mais nous voulons montrer ici en effet qu'un corpus n'est pas un réceptacle mais une matrice. Combien il n'est pas un lieu de validation mais un lieu d'invention. Comment il n'est pas l'*observatoire* froid d'un savoir linguistique *a priori*, mais un *observé* vivant, dynamique dont la description doit aboutir à une connaissance linguistique *a posteriori*, à des modèles sémantiques à découvrir, à des typologies textuelles à inventer. Il y a là une nuance importante dans le procès de la recherche qui n'est pas sans rappeler la différence entre déduction et induction. « *C'est finalement le corpus qui fait la théorie* » n'hésite pas à souligner ici même en 2002 [Dalbera 2002 : 99], mettant en bon ordre le sens de la démarche.
- 13 Constitué certes sur la base d'hypothèses de travail conscientes et contraignantes (voir notamment très précisément [Pincemin 1999]), le corpus devient *l'objet* de la recherche et non *l'outil* de la recherche. Il n'est pas l'instrument du linguiste mais sa main et sa chose. Il n'est pas reflet d'un fait linguistique transcendant ni l'échantillon d'une langue déjà-là, mais il est *le* fait linguistique en lui-même que l'on juge digne d'être observé dans son immanence.

- 14 Parce que le sens (politique) n'est jamais donné, le corpus ne peut-être un recueil de données, à fonction documentaire. Encore moins peut-on le définir comme un recueil d'exemples : le sens (politique) ne se laisse pas exemplifier. Pas davantage comme une chambre froide : le sens est toujours brûlant. Jamais comme un sac de mots : le sens ne s'achète pas en vrac. Le corpus est pour nous, en lui-même, une quête vers le sens qu'il construit. Il est « *médiation* » nécessaire selon le joli mot de [Mellet 2002 : 9] c'est-à-dire, au sens philosophique, un *processus créateur* indispensable pour le linguiste dans la pratique de sa science.
- 15 Tels sont quelques postulats que nous reprenons, ici, du côté du discours politique, à la *linguistique de corpus* avant que l'expression ne se galvaude et que les corpus ne soient revendiqués par chacun<sup>8</sup>.

Le corpus et son extérieur : la tension disciplinaire

- 16 Les corpus politiques – contrairement aux corpus grammaticaux, phonologiques ou lexicaux, et encore, à un autre niveau, aux corpus littéraires – sont tout entier référentiels : c'est dans leur nature, leur vocation, leur raison d'être, de renvoyer au monde réel. En cela, ils posent dans leur clôture – un corpus est toujours clos même si sa clôture est provisoire – immanquablement la question du rapport entre leur intérieur et leur extérieur, réduction modeste et technique, sans doute, du rapport entre la langue et le monde.
- 17 De quoi est fait l'intérieur d'un corpus politique ? Sur quels critères extérieurs à la langue, les textes ont-ils été qualifiés de politiques ? Quels sont les relations qu'entretiennent texte, intertexte, contexte, extra-texte dans la linguistique de corpus ? Un corpus est-il une entité sémantique auto-suffisante ou n'a-t-on pas toujours recours à des ressources extérieures intertextuelles et étrangères extra-textuelles pour l'interpréter ? Dès lors qu'il se suffirait à lui-même, de quoi pourrait témoigner un corpus au-delà de lui-même ? Où commencent et où s'arrêtent nos capacités d'inférences sociolinguistiques ? *Quid* du *bond interprétatif* noté par tous les auteurs au moment de la compréhension des textes, bond d'autant plus crucial que l'on admet, précisément, que le sens n'est jamais qu'interprétation ? Etc. Ces questions, ressassées depuis l'origine par l'AD politique, mettent en réalité à jour des tensions disciplinaires importantes, jamais retombées, entre la linguistique et les sciences humaines [par exemple : Robin 1973 et 1986]. A défaut de résoudre un jour le problème visiblement insoluble de l'interdisciplinarité en général et de l'interdisciplinarité des sciences du langage en particulier [voir néanmoins *Cahiers Ferdinand de Saussure* 2001], il faut s'appliquer à soulager ces tensions en définissant la place de chacun et le rôle essentiel, primordial, nécessaire (quoique non-suffisant) de la linguistique dans l'analyse des corpus textuels.
- 18 Il convient en effet d'abord de rappeler l'évidence : la chair du corpus est le verbe. L'intérieur du corpus est composé de textes, de phrases, de mots, de syllabes, de lettres c'est-à-dire d'un matériel linguistique. Quand bien même les textes auront été sélectionnés et organisés en corpus sur la base d'hypothèses de travail extra-linguistiques, la matérialité textuelle du corpus – après la *matérialité discursive* du célèbre colloque de 1981 – est un constat, pour nous, indépassable. En cela, le traitement linguistique n'est pas négociable même pour les disciplines qui défient la linguistique. A moins de prétendre connaître le sens des textes avant de les avoir dûment – *i.e.* linguistiquement – analysés, à moins de croire que les textes sont transparents et que le sens est immédiat, l'historien, le sociologue, le psychologue ne peut faire l'économie de la linguistique comme *science des textes*, dans l'appréhension du corpus. Pour [Mazière 2005]

qui en retrace l'historique, l'école française d'Analyse du Discours, dont les objectifs étaient pourtant très idéologiques, se caractérise par cette primauté accordée à la linguistique, considérée, ici à juste titre, comme science pilote. C'est cette primauté qui construit les ponts entre l'AD politique constituée dans les années 1970 et la linguistique de corpus politiques aujourd'hui : le corpus se présente sous une forme matérielle qui appartient bien à la langue et à laquelle l'analyste ne peut se soustraire.

- 19 Ainsi, dans ce numéro, tous les contributeurs, quels que soient leur rattachement institutionnel et leur appartenance disciplinaire, prennent une posture de linguiste face à l'intérieur du corpus, et ont recours aux sciences du langage dans le protocole méthodologique qu'ils mettent en place face aux textes. La question de la méthode est évidemment cruciale dans un examen scientifique mais elle prend un tour particulier pour nous car l'objet (le corpus textuel), une bonne partie des ressources interprétatives (l'intertexte ou le contexte) et encore le métalangage scientifique pour traiter l'objet sont de même nature. En termes plus rapides : le texte et sa glose, c'est-à-dire encore, d'une certaine manière, l'intérieur du corpus et une partie majeure de son extérieur sont du langage, avec un risque de confusion évident. C'est pour cette raison que l'analyse du discours ou la linguistique de corpus insiste sur le *décentrement* nécessaire du chercheur – qui doit se comporter en analyste et non en locuteur natif – par rapport à son objet d'étude, et sur la recherche de *méthodes* qui permettent l'*objectivation* de la lecture, afin que celle-ci cesse d'être « naturelle » pour être scientifique<sup>9</sup>.
- 20 Dans ce cadre, il est logique de retrouver en bonne place dans ce numéro, la linguistique quantitative assistée – faut-il encore le préciser aujourd'hui ? – par ordinateur. Quatre articles y sont consacrés. Cela atteste de la vitalité nouvelle du domaine avec le succès bi-annuel des Journées internationales d'Analyse statistique de Données Textuelles et le dynamisme des revues en ligne française et canadienne *Lexicométrie* et *Astrolabe* qui lui sont entièrement ou en partie dédiées. Après d'autres, nous avons montré récemment [Mayaffre 2005] que la logométrie et l'ordinateur restent effectivement le moyen le plus performant pour embrasser de grandes masses textuelles politiques. La médiation informatique et l'approche algo-rithmique des textes permettent de désubjectiver la lecture comme l'avait réclamé [Pêcheux 1969] dès la fin des années 1960. Cela permet de tendre vers une *herméneutique numérique* [Mayaffre 2002-a] dont la caractéristique première tient dans la précision de la description et le contrôle de l'interprétation grâce à une déconstruction / reconstruction du texte et une lecture systématique et exhaustive des corpus (si possible *réflexifs* [Mayaffre 2002-b]). Ici même, en fin de numéro, le compte rendu de l'ouvrage de [Mazière 2005] nous permet de revenir sur une méthode que les articles de Arnold, Desmarchelier, Labbé, Labbé et Monière, Leblanc et Martinez illustrent abondamment.
- 21 A côté de l'approche quantitative, la rhétorique traditionnelle avec des concepts éprouvés – avant d'être oubliés – depuis Aristote est mobilisée par Carine Duteil. Mais comme le souligne l'auteure, cette mobilisation se fait non du point de vue des sciences de l'information et de la communication mais « dans le champ de la linguistique » [Duteil 2005] et plus précisément dans celui de la sémantique interprétative : c'est en s'appuyant sur des traits linguistiques formalisés du corpus (traits qu'elle se propose d'ailleurs de baliser en corpus pour plus encore de formalisme), que l'auteure avance dans son analyse des textes.
- 22 Le souci méthodologique, du côté de la linguistique, est plus poussé peut-être chez Laurent Rouveyrol puisque c'est, précisément, une conciliation entre méthode

quantitative et méthode qualitative, que l'auteur propose dans la perspective d'une logométrie intégrative. Cette conciliation, encore expérimentale, doit croiser une perception macro et un traitement micro des traits sémantiques des discours : elle apparaît effectivement possible du moment que l'analyse s'appuie, toujours, sur la matérialité discursive du corpus et utilise la médiation informatique pour assurer une homogénéité – pour ne pas dire « neutralité » – du traitement.

- 23 Bref, la méthode est pour nous importante car elle établit un sas entre l'intérieur formel et l'extérieur informel du corpus, entre l'objet et le sujet, entre la langue et le monde. En ce sens nous avons plaidé ailleurs [Mayaffre 2005] pour une démarche non pas idéale mais contrôlée qui isolerait très clairement l'élaboration des hypothèses de travail qui président à la construction du corpus (hypothèses pas nécessairement linguistiques d'ailleurs), le traitement linguistique de l'objet corpus (traitement nécessairement linguistique quant à lui) et le moment de l'interprétation des textes durant lequel l'ensemble de nos compétences culturelles sont mobilisées. Car, ne le cachons pas, l'analyse des corpus politiques nous projette tôt ou tard hors du corpus et hors de la linguistique, vers l'histoire, la sociologie, les sciences politiques. Les ressources interprétatives ne sont évidemment pas toutes textuelles et ne sauraient entièrement se trouver consignées dans le corpus. Pour tout dire, notre référent est hors corpus, debout dans une réalité sociale à propos de laquelle le linguiste, seul, n'en peut mais. Pilote, la linguistique n'est donc pas suffisante pour traiter des corpus politiques, quand bien même il doit lui revenir le rôle majeur d'analyser les textes.
- 24 Sans doute est-ce pour cette raison que les analyses textuelles, tout en se revendiquant de la linguistique, entendent participer, au-delà d'elle, au projet plus global d'une « sémiotique des cultures » [Rastier 2001, *Épilogue* : 275-284] dans laquelle l'apport d'autres disciplines est évident (même s'il reste à définir). Mieux : c'est précisément en cela que la linguistique de corpus (politiques) reprend le projet saussurien - le vrai projet des manuscrits [Saussure, *sine data*-1996 et 2001], non celui, sur-interprété du *Cours de linguistique générale*. La linguistique en effet n'a pas seulement comme but fermé l'introspection du système. Elle a surtout l'objectif général, ouvert et ambitieux, d'extraire de la culture humaine son essence langagière, afin de mieux la comprendre. Dans cette tâche, selon [Bouquet 2005] qui se fait porte-parole de son fondateur, la linguistique aussi souveraine soit-elle (ici, dans l'analyse des corpus textuels) « ne saurait, contrairement à sa tendance actuelle, être le propre juge de ses résultats ». Tout au contraire, « on peut tenir (...), dans la perspective ouverte par *L'essence double du langage* [Saussure, *sine data*-1996 et 2001], qu'elle ne se justifie que des résultats qu'est susceptible de produire son dialogue avec les autres sciences humaines et sociales. »
- 25 En d'autres termes, le corpus de textes politiques est un objet linguistique à part entière mais sous la lumière d'un éclairage extérieur :
- « Pour aborder sainement la linguistique, il faut l'aborder du dehors, mais non sans quelque expérience des phénomènes prestigieux du dedans<sup>10</sup>. Un linguiste qui n'est que linguiste est dans l'impossibilité à ce que je crois de trouver la voie permettant de classer les faits » [Saussure, *Note Item*, cité par Adam 2001 : 201].

## BIBLIOGRAPHIE

- Adam J.-M. (1990). *Eléments de linguistique textuelle*. Bruxelles : Mardaga.
- Adam J.-M. (1999). *Linguistique textuelle. Des genres de discours aux textes*. Paris : Nathan.
- Adam J.-M. (2001). « Discours et interdisciplinarité », *Cahiers Ferdinand de Saussure* 54 : 201-218.
- Aijmer B. & Altenberg K. (éd.) (2002). *Advances in Corpus Linguistics*. New York : Rodopi.
- Amossy R. (éd.) (2002). *Pragmatique et analyse des textes*. Tel-Aviv : Presses de l'Université de Tel-Aviv.
- Biber D. (1988). *Variation across speech and writing*. Cambridge : Cambridge University Press.
- Biber D. (1995). *Dimensions of Register Variation : A Cross-linguistic Comparison*. Cambridge : Cambridge University Press.
- Biber D., Conrad S. & Reppen R. (1998). *Corpus linguistics. Investigating language, Structure and Use*. Cambridge : Cambridge University Press.
- Bouquet S. (1997). *Introduction à la lecture de Saussure*. Paris : Payot.
- Bouquet S. (2005). « Après un siècle, les manuscrits de Saussure reviennent bouleverser la linguistique », *Texte !* <[http://www.revue-texto.net/Saussure/Sur\\_Saussure/Bouquet\\_Apres.html](http://www.revue-texto.net/Saussure/Sur_Saussure/Bouquet_Apres.html)>.
- Bronckart J.-P. (1997). *Activité langagière, textes et discours*. Lausanne-Paris : Delachaux et Niestlé.
- Brunet É. (1981). *Le Vocabulaire français de 1789 à nos jours*. Paris-Genève : Champion-Slatkine.
- CAHIERS FERDINAND DE SAUSSURE (2001). « Actes du colloque international : Ferdinand de Saussure et l'interdisciplinarité des sciences du langage » (novembre 1999), 54 : 144-338.
- Combettes B. (1983). *Pour une grammaire textuelle*. Bruxelles : De Boeck-Duculot.
- CORPUS (2002). « Corpus et recherches linguistiques » n°1, numéro coordonné par Sylvie Mellet, 175 pages. (Consultable sur <http://revel.unice.fr/corpus/>).
- Dalbera J.-Ph. (2002). « Le corpus entre données, analyse et théorie », *Corpus* 1 : 89-105.
- Dubois J. & Sumpf J. (1969). « Problèmes de l'analyse du discours », *Langages* 13 : 3-7.
- Duteil-Mougel C. (2005). « Ethos, logos et pathos dans les textes politiques contemporains », in P. Marillaud & R. Gauthier (éds.) *Rhétorique des discours politiques*, Actes du 25<sup>ème</sup> Colloque d'Albi « Langages et Significations ». Toulouse : CALS/CPST, pp. 343-355.
- Guilhaumou J. et alii (1994). *Discours et archive*. Liège : Mardaga.
- Habert B., Nazarenko A. & Salem A. (1997). *Les linguistiques de corpus*. Paris : Colin.
- Hjelmslev L. (1943) (1968-1971). *Prolégomènes à une théorie du langage*. Paris : Minuit.
- Halliday M. A. K. & Hasan R. (1976). *Cohesion in English*. Londres : Longman
- Vergely P. (éd.) (2005). *Rôle et place des corpus en linguistique*, actes du colloque JETOU'2005. Toulouse : ERSS



- Mayaffre D. (2002-a). « L'Herméneutique numérique », *L'Astrolabe. Recherche littéraire et Informatique*, (<http://www.uottawa.ca/academic/arts/astrolabe/>).
- Mayaffre D. (2002-b). « Les corpus réflexifs : entre architextualité et hypertextualité », *Corpus 1* : 51-69.
- Mayaffre D. (2005). « Analyse du discours politique et logométrie. Point de vue pratique et théorique », *Langage et Société* 114 : 91-111.
- Mazière F. (2005). *L'analyse du discours. Histoire et pratiques*. Paris : PUF.
- Mellet S. (2002). « Corpus et recherches linguistiques : introduction », *Corpus 1* : 5-13.
- Pêcheux M. (1969). *Analyse automatique du discours*. Paris : Dunod.
- Peschanski D. (éd.) (1984). *Histoire moderne et contemporaine et informatique*, n°4.
- Pincemin B. (1999). « Construire et utiliser un corpus : le point de vue d'une sémantique textuelle interprétative », in A. Condamines et alii (éds.), *Corpus et traitement automatique des langues : pour une réflexion méthodologique*, Actes de l'atelier thématique TALN. Cargèse, pp. 26-36.
- Rastier F. (1989). *Sens et textualité*. Paris : Hachette.
- Rastier F. (2001). *Arts et sciences du texte*. Paris : PUF.
- Rastier F. (2005-a). « Enjeux épistémologiques de la linguistique de corpus », in G. Williams (éd.), *La linguistique de corpus*. Rennes : PUR, pp. 31-45. [ En ligne sur *Texte !* ([http://www.revue-texto.net/Inedits/Rastier/Rastier\\_Enjeux.html](http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html))]
- Rastier F (2005-b). « Discours et texte » (première partie), *Texte !* ([http://www.revue-texto.net/Reperes/Themes/Rastier\\_Discours.html](http://www.revue-texto.net/Reperes/Themes/Rastier_Discours.html)).
- Robin R. (1973). *Histoire et Linguistique*. Paris : A. Colin.
- Robin R. (1986). « Histoire et Linguistique : le malentendu continue », *Langages* 81 : 121-128.
- Saussure F. de (s.d. – 1996). *De l'essence double du langage* (transcription diplomatique établie par Rudolf Engler d'après le manuscrit déposé à la Bibliothèque de Genève). en ligne sur *Texte !*, décembre 2004 - juin 2005 (<http://www.revue-texto.net/>).
- Saussure F. de (s.d. – 2001). *Ecrits de linguistique générale* [Simon Bouquet et Rudolf Engler (éds)]. Paris : Gallimard.
- Scheer T. (2004-a). « Présentation du volume. En quoi la phonologie est vraiment différente », *Corpus 3* : 5-85.
- Scheer T. (2004-b). « Le corpus heuristique : un outil qui montre mais ne démontre pas », *Corpus 3* : 153-193.
- Tognini-Bonelli E.(2001). *Corpus Linguistics at work*. Amsterdam-Philadelphia : John Benjamins Publishing Company.
- Van Dijk T. (1984). « Texte », dans *Dictionnaire des littératures de langue française*, Beaumarchais et alii (éds.). Paris : Bordas.
- Williams G. (éd.) (2005). *La linguistique de corpus*. Rennes : PUR.

## NOTES

- 1.. C'est précisément au nom de cet historique complet que le lecteur pourra consulter sur papier ou en ligne (<http://revel.unice.fr/corpus/>) que nous nous permettons ici de ramener plusieurs décennies d'histoire et une bibliographie foisonnante aux quatre seules références données plus haut.
- 2.. On l'aura compris, nous prenons le terme *contexte* dans son sens englobant et jouons avec toute son étendue. Le contexte c'est d'abord le texte, c'est ensuite l'intertexte ou ce que d'aucuns ont appelé le *co-texte*, c'est enfin le contexte extra-linguistique c'est-à-dire la situation d'énonciation, les conditions historiques de production du discours, etc.
- 3.. De manière plus technique, notons que la statistique linguistique, qui s'est imposée comme une méthode performante pour traiter des grands corpus politiques, a toujours considéré, à l'intérieur même du traitement mathématique, les populations particulières (les mots, les segments, les codes) par rapport à l'ensemble textuel que représente le corpus. L'unité mathématique, aussi, prend du sens par rapport au tout.
- 4.. On trouvera plusieurs formulations pour exprimer cette idée sous la plume de Rastier : « le texte est pour une linguistique évoluée l'unité *minimale* » [Rastier 2005-a : 31, la casse est de l'auteur] ; « le texte est l'unité fondamentale... » [Rastier 2001 : 108]. On rapprochera ces formulations de celle, définitive, de [Adam 2001 : 216] : « Si, comme le dit Saussure, la langue n'est créée qu'en vue du discours, la linguistique a non seulement pour objet empirique mais pour objet théorique cette unité de communication-interaction langagière qu'on appelle un TEXTE (ou un DISCOURS)... » (la casse est de l'auteur).
- 5.. Ultime : de la lettre au mot, du mot à la phrase, de la phrase au texte, du texte au corpus textuel... L'expansion de la linguistique semble devoir s'arrêter là. Ce qui se passe au-delà du corpus est, par définition, hors de portée du linguiste.
- 6.. Ainsi Chomsky peut-il encore déclarer en 1999 : « Corpus linguistics does not exist » (cité par [Rastier 2005 : 40] qui y voit un bel exemple de dénégation).
- 7.. Et cette différence de point de vue puise sans doute son origine dans la nature lexicale (donc finie) vs. textuelle (donc infinie) des corpus en phonologie vs. en sémantique.
- 8.. Une fièvre a en effet saisi la communauté linguistique en France ces derniers mois. Plus d'une demi-douzaine de journées, tables rondes, colloques a été consacrée au(x) corpus en 2004-2005. De manière significative, les jeunes chercheurs parisiens et toulousains ont organisé à quelques mois d'intervalles des rencontres sur le sujet ([Coldoc'2005 : actes non parus] et [Jetou'2005 : actes sous presse]). Plus que jamais, la revue CORPUS justifie sa place éditoriale et doit organiser la réflexion épistémologique, aujourd'hui foisonnante, sur la place des corpus en linguistique.
- 9.. Ajoutons aussi de manière plus prosaïque que ce souci d'objectivité est doublé par le caractère souvent brûlant des textes politiques. Le chercheur doit se comporter en analyste, non en locuteur, nous l'avons dit, et moins encore en citoyen.
- 10.. Derrière la modestie de la formulation (« quelque expérience »), cette première phrase établit bien le rôle indispensable et *nécessaire* (« non sans... ») de la linguistique ; la seconde phrase établissant son caractère *non-suffisant*.

---

AUTEUR

**DAMON MAYAFFRE**

CNRS, BCL, Nice